PCT/GB 2004 / 0 0 4 1 4 8

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 19 OCT 2004
WIPO PCT

PRIORITY DOCUMENT

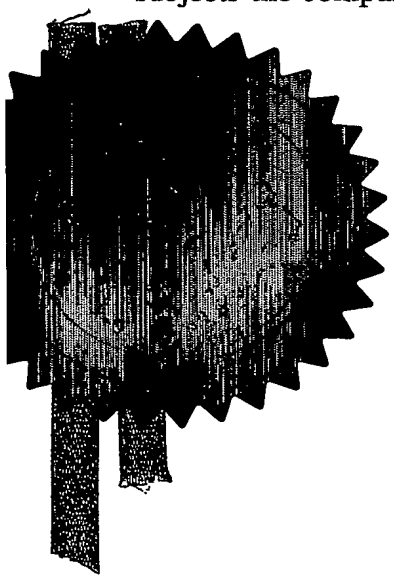
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed Stephen Hordley

Dated 5 October 2004

Patents Form 1/77

The Patent Office

1/77

2 9 SEP 2003

Act 1977

THE PATENT OFFICE
2 9 SEP 2003
LONDON

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP10 8QQ

1. Your reference

30SEP03 E840770-1 D03052
P01/7700 0.00-0322763.4

A30295

2. Patent application number
(The Patent Office will fill in this part)

0322763.4

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

BRITISH TELECOMMUNICATIONS public limited company
81 NEWGATE STREET
LONDON, EC1A 7AJ, England
Registered in England: 1800000

Patents ADP number *(if you know it)*

1867002✓

If the applicant is a corporate body, give the country/state of its incorporation

UNITED KINGDOM

4. Title of the invention

TIME-SLOT ASSIGNMENT PROCESS

5. Name of your agent *(if you have one)*

"Address for Service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

BT GROUP LEGAL
INTELLECTUAL PROPERTY DEPARTMENT
HOLBORN CENTRE
120 HOLBORN
LONDON, EC1N 2TE

Patents ADP number *(if you know it)*

~~**1867001**~~ 1867007

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day/month/year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.

(See note (d))

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-
Description	18
Claim(s)	3
Abstract	1
Drawing(s)	7 + 7 done

10. If you are also filing any of the following, state how may against each item

Priority Documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	YES
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	2 COMPLETE SETS OF MULTI-STAGE MATCHING PROCESS

11. I/We request the grant of a patent on the basis of this application.

Signature(s) Date:

29 September 2003

LIDBETTER, Tomothy Guy Edwin, Authorised Signatory

12. Name and daytime telephone number of person to contact in the United Kingdom — Rod HILLEN — 020 7492 8140

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

TIME-SLOT ASSIGNMENT PROCESS

This invention relates to a time-slot assignment process for use in a contention resolution scheme for switches and switch arrangements in a communications network. In particular, but not exclusively, the invention relates to an aggregated channel assignment technique for multi-stage buffering and switching in cell/packet switches and switching arrangements in a communications network.

The aggregated channel assignment solution is suitable for use with scalable matching algorithm which matches asymmetric request matrices for switches and switch arrangements as described in our co-pending patent application filed herewith whose contents are hereby incorporated by reference, entitled "MULTI-STAGE MATCHING PROCESS".

This invention complements the multi-stage matching invention described in the inventors co-pending patent application entitled "MULTISTAGE MATCHINGPROCESS", filed herewith as both inventions provide means for reducing computing steps in frame-based scheduling. Nonetheless, it is not a requirement that this invention be used in conjunction with multi-stage matching, merely an option.

Single stage switch/network fabrics are well known in the art. Time-slot assignment of cell or packet requests between the input and output ports of switches and networks, such as is shown schematically in Figure 1, is conventionally performed in a single stage. The number of input and output ports is taken here to be LN. If the switch fabric is a single-stage switch matrix, whether implemented as a space switch or as a wavelength switch, and cells or packets are switched between the input and output ports in a frame with a duration of F fixed-length time slots, then the time slots can be chosen using a path-search algorithm in a conventional three-stage TST switching network. Each input and output port of the 3-stage logical switch represents a time slot occupied by an individual cell or packet. There are LNF such input and output ports.

When each "connection" path between the LNF input and output ports across the 3-stage switch is implemented sequentially, the computing time (steps) is $O(LNF^2)$. If we assume, for example, that the computing time allowed for time-slot assignment is as much as a full frame of F time slots, i.e. $F\tau$ seconds, this requires a single processor with a clock speed $O(LNF^2/F\tau) = O(LNF/\tau)$ Hz. For a 10 Tbit/s cell/packet switch with $LN = 1024$ ports at $B = 10$ Gbit/s line rate, $F = LN$ time slots per frame and $\tau = 50$ nsec, this single processor would need a clock speed of 20 THz, which is twice the capacity of the switch it is controlling. A single sequential processor at 20 THz is impossible. More precisely, the ratio of total processing capacity to switch capacity is

$$\frac{Processing Capacity}{Switch Capacity} = \frac{O(LNF/\tau)}{LNB} = O\left(\frac{F}{\tau B}\right) = O\left(\frac{F}{500}\right) \qquad \text{...eqtn. 1}$$

which is a ratio of $1024/500 = 2$ for this example.

The inventors earlier patent specification WO 01/67802 entitled "Packet Switching" describes a doubly parallel implementation of the path-search algorithm, in order to reduce the computing steps to $O(LN\log_2 F)$. This employed LN parallel processors, each one of which contained a compact superconcentrator for finding and seizing in parallel the required number of free, common time slots between the occupancy vectors of a pair of input and output ports. Each of the LN parallel processors must have a clock speed of at least $O(LN\log_2 F/F\tau)$ Hz. In the same example, this is 200 MHz, which in itself is quite acceptable. But each compact superconcentrator possesses an omega network with O(F) ports. So the total "throughput" of all the compact superconcentrators is O(LNFx200) MHz = O(200 THz). This is 20 times the capacity of the switch being controlled, which is absurd. More precisely:

$$\frac{Processor Throughput}{Switch Capacity} = \frac{O\left(\frac{L^2N^2F\log_2 F}{F\tau}\right)}{LNB} = O\left(\frac{LN\log_2 F}{\tau B}\right) = O\left(\frac{LN\log_2 F}{500}\right) \qquad \text{...eqtn. 2}$$

which is a ratio of 20 for this example.

A large reduction in processing capacity can be achieved by using only a single level of parallelism from WO 01/67802. LN port processors would still operate in parallel, each one finding the required number of free, common time slots between a pair of input and output ports, but the time slots themselves would be found by sequential

inspection. All input-output port pairs are processed after LN steps of parallel processing. There are therefore LN processors, each taking up to O(LNF) computing steps altogether. Hence

$$\frac{ProcessorCapacity}{SwitchCapacity} = \frac{O\left(\frac{L^2N^2F}{F\tau}\right)}{LNB} = O\left(\frac{LN}{\tau B}\right) = O\left(\frac{LN}{500}\right) \qquad \text{...eqtn. 3}$$

This ratio is around 2 again, as in eqtn1. The processing capacity is still greater than the switch capacity being controlled. Furthermore, each sequential processor must have a clock speed $O(LN/\tau)$ = 20 GHz, which is out of the question today.

For such large switches, the processing capacities required to assign time slots across a single-stage switch fabric, using existing time-slot assignment algorithms and circuits, are too large. Without improved algorithms and circuits, multi-stage switch fabrics must be employed.

Although multi-stage switches and networks must now be considered as a means of reducing processing complexity, it is insufficient simply to change the single-stage switch matrix of Figure 1 into a 3-stage space switch fabric. Together with the outer time switches, the structure becomes a logical, 5-stage TSSST switch. Of course if agile wavelength channels are being used to implement the space switching, this can be a logical, 5-stage TλSλT switch, for example. However, all the processing requirements described above would still be needed to assign the time slots across the 3-stage space switch. In addition, further processing would be needed to establish the paths through the 3-stage space switch fabric in every time slot.

Simply applying the same number of time slots to a multi-stage switch fabric in this way is of no benefit in reducing the processing complexities needed for channel assignment.

The invention seeks to mitigate and/or obviate the disadvantages of known time-slot assignment techniques by providing a multi-stage time-slot assignment process for a scheduling process in which:

i) time slots are aggregated from multiple ports to provide a common pool of time-space channels; and

ii) the 1st, 3rd, 5th and optionally 7th stages provide time-slot interchange capability, as decried herein below.

A first aspect of the invention seeks to provide a multi-stage time-slot assignment process for an input-queued switch arrangement in a communications network, the switch comprising a plurality of N ingress elements and N egress elements, each of the ingress elements having a number L of ingress subelements and each of the egress elements having a plurality L of egress subelements, the switch arrangement being arranged to have L or more real middle stage space switches of size N x N, and having F or more time-slots, the time-slot assignment process between the said ingress subelements and egress sub-elements comprising the steps of: aggregating F time slots from each of a plurality L in number of said ingress subelements to form an ingress element having a plurality LF or more in number of time-space channels which are pooled between the L subelements of each ingress element and the L subelements of each egress element, wherein each time-space channel corresponds to a different logical middle-stage switch of the packet switch arrangement so that the number of logical input elements and logical output elements for which channel assignment is performed through the middle stage of the switch is N; performing time-space channel assignment through the middle stage of the switch between the logical input elements and the logical output elements; providing time-slot interchange capabilities at the 1st, 3rd, 5th stages; and performing time-slot assignment between the ingress subelements of the ingress elements and the logical output ports of the 3rd stage switches, and performing time-slot interchange between the logical input ports of the 5th stage switches and the egress subelements of the egress elements.

The total number of ingress sub-elements in each element may not equal the number of egress subelements in each element, providing the total number of ingress subelements for the switch equals the total number of egress subelements for the switch.

The switch arrangement may be arranged to switch cells or packets. Alternatively, the switch arrangement may be a circuit switch.

The switch arrangement may a cell switch arrangement capable of switching packets.

The time-slot assignment process according to the first aspect may further comprise a 7th stage providing a time-slot interchange capability.

The sub-elements may comprise ports, and the elements comprise aggregations of ports.

The switch arrangement may be an input-queued switch comprising VOQs which are implemented in random access memory RAM (for example, when the switch arrangement is arranged to switch cells or packets).

The time-slot assignment process according to the first aspect may further comprise the recursive decomposition of the three stages in each element into seven stages using the steps according to the first aspect.

A second aspect of the invention seeks to provide a switch arrangement in which time-slots are assigned by a time-slot assignment process according to the first aspect.

A third aspect of the invention seeks to provide a network having a switch arrangement according to the second aspect.

A fourth aspect of the invention seeks to provide a suite of at least one computer program arranged to implement the time-slot assignment process according to the firs aspect.

The suite of at least one computer program may be at least partly arranged to be implemented in hardware. The suite of at least one computer program may be at

least partly arranged to be implemented in software. Thus it is possible in some embodiments of the invention to implement the time-slot assignment process using a combination of hardware components and software.

Those skilled in the art will appreciate that the preferred features of the invention as defined by the dependent claims may be combined with each other where appropriate any suitable manner apparent to those skilled in the art and with any of the aspects where appropriate and in any suitable manner apparent to those skilled in the art.

The embodiments of the invention will now be described by way of example with reference to the accompanying drawings in which:

Figure 1 shows schematically input and output ports of a cell/packet switch or cell/packet switching network;

Figure 2 is a schematic diagram provides a 7-stage switching architecture for aggregating the time slots of L ports into a common pool of LF channels;

Figure 3 shows schematically the Assignment of cells or packets to logical inputs and outputs of the 7-stage logical switch architecture in an embodiment of the invention;

Figure 4 shows schematically paths and cell/packet assignments after the first path-search phase through the middle-stage switches in the embodiment of the invention shown in Figure 3;

Figure 5 shows schematically paths and cell/packet assignments after the second path-search phase through the individual sub-networks in the embodiment of the invention shown in Figure 3;

Figure 6 shows schematically paths assigned to each of the 16 cells or packets through the 7-stage logical switching network in the embodiment of the invention shown in Figure 3; and

Figure7 shows schematically re-ordering of cells or packets to prevent mis-sequencing when the 7th-stage time-slot interchangers are left out in the embodiment of the invention shown in Figure 3.

The best mode of the invention as currently contemplated by the inventors will now be described with reference to the accompanying drawings. Those skilled in the art will recognise that the detailed description provided of specific embodiments of the invention below is not intended to represent the only means of implementing the invention, and that any suitable means which is apparent to a person skilled in the art may be used where appropriate to implement any specific feature of the invention as described.

This invention relates to the channel assignment part of a frame-based scheduling algorithm. However, it is also widely applicable to the general problems of time-slot and channel assignment in switches. The technique provides a means for reducing the computing steps required to implement the parallel Clos path-searching algorithm, as applied to the time-slot assignment of cells and packets in frame-based scheduling (and as described by the inventors in International PCT patent application WO 01/67802 entitled "PACKET SWITCHING", the contents of which are hereby incorporated by reference). This is achieved by defining a larger set of channels than there are time slots in the frame and selecting (path-searching) those channels through the middle stage of a multi-stage switching architecture. The outer stages of this architecture are themselves then recursively expanded into 3-stage switches and their time slots chosen (path-searched). The result of this is to require time-slot interchangers to be inserted as additional switching stages within a multi-stage switching fabric.

This principle enables channel assignment to be performed within and between smaller switches or sub-networks, hence reducing the computing complexity. The principle can be applied to sub-networks in the form of individual smaller crossbars, individual smaller cell/packet switches, individual rings and individual PONs. When this principle is applied to multi-stage fabrics of individual smaller crossbars, the

switching fabric (i.e. the space switches), excluding the additional time-slot interchangers, remains the same as would be required without this technique, i.e. there is no speed-up required of the space switch fabric to implement the technique, just additional buffers/time-slot interchangers. These could of course be implemented in the form of additional cell/packet switch linecards. If both contiguity and no mis-sequencing of cells/packets from the same VOQ are required at the output of the cell/packet switch, then 3 additional stages of time-slot interchangers (buffers) are necessary. But if only no mis-sequencing is required, allowing cells/packets from the same VOQ to exit the cell/packet switch with cells/packets from other VOQs interspersed between them, then only 2 additional stages of time-slot interchangers (buffers) are required. These requirements compare favourably with the 2-stage Birkhoff-von Neumann switch, which requires two space switch fabrics and 1 additional stage of buffers, and even more favourably with the parallel packet switch architecture, which can require a fabric speed-up of 3.

Although this description applies decomposition only once to the sub-networks, turning them into 3-stage TST switches and the initial 3-stage logical switch into a 7-stage logical switch, the principle could of course be applied recursively again to the resulting 3-stageTST switches within each sub-network, turning them into 7-stage logical switches and the initial 3-stage logical switch into a 15-stage logical switch, and so on.

The inventors have found that in order to reduce the processing complexity needed for channel assignment, the number of channels to be assigned is increased. This is achieved by grouping together L ports into sub-groups or sub-networks, so that their time slots are aggregated into a common pool of LF time-space channels. See Figure 1 of the accompanying drawings. Each channel corresponds to a different middle-stage space switch (both real and logical). There are L real middle-stage space switches, each of size NxN and each of which exists in F time slots. This aggregation technique has the effect of reducing the number of logical ports for which channel assignment is performed through the middle stage of the switch to just N. Since the number of computing steps (see eqtns.1-3) is generally related to the number of ports, this reduces the processing complexity.

Figure 2 is a 7-stage switching architecture. The sub-groups or sub-networks of ports are themselves recursively decomposed into 3-stage switches. This is necessary because two cells or packets sent or received by any of the LN real switch ports may be assigned to the same time slot through different middle-stage switches. Since real ports have only one transmitter and receiver, they cannot transmit or receive more than one cell or packet in the same time slot. Their cells or packets must be transmitted in different time slots within the sub-group or sub-network of ports to other, different real ports (or linecards), one of which could be the same as the transmitting one, where they must be time-slot interchanged as required. The 3-stage sub-networks therefore consist in an LxL space switch (or wavelength switch, for example) which is configured in F different time slots. Thus the 1st, 3rd, 5th and 7th stages of the 7-stage architecture are time-slot interchangers. The switches in the 1st and 7th stages represent the real ports or linecards of the overall switch. The 1st-stage time-slot interchangers in reality represent the forwarding of each cell or packet from the correct VOQ in the correct time slot.

Of course there is nothing new about the recursive construction that decomposes the outer switches of a 3-stage switch into 3-stage switches themselves, resulting in 7 stages altogether. This was evident from the earliest work of Clos to the explicit statement in United States patent number US 5,801,641.

The invention provides a time-slot assignment process for a scheduling process in which:

iii) time slots are aggregated from multiple ports to provide a common pool of time-space channels; and

iv) the 1st, 3rd, 5th and optionally 7th stages provide time-slot interchange capability.

The 7th stage does not need to be implemented in certain circumstances.

Time-slot interchangers in the 1st stage represent the forwarding of each cell or packet from their positions in their VOQs into the switch fabric in the correct time slot.

Before channel assignment can be implemented, it is first necessary to assign the LNF inputs and LNF outputs of the network to the cells or packets to be switched in the frame. This does not need to be done for every cell or packet individually. Instead, on the input side of the switch, every real port or linecard would assign a number of contiguous inputs (i.e. time slots) to the $a_{i,j}$ successfully accepted requests of each of its LN VOQs, i.e. from VOQ $a_{i,1}$ to $a_{i,LN}$, obtained from a previous matching phase. Similarly, on the output side, every real port or linecard j would assign a number of contiguous outputs to the $a_{i,j}$ successfully accepted requests from each of the LN inputs, i.e. from VOQ $a_{1,j}$ to $a_{LN,j}$.

If contiguity of outgoing cells or packets from the same VOQ is required, then all 4 stages of time-slot interchanger are required. But if contiguity of outgoing cells or packets from the same VOQ is not required, the last stage of time-slot interchange can be left out; i.e. the real output ports would have no buffers and cells or packets would leave the switch in the time slots assigned to them within the sub-group or sub-network (stages 5 to 7). However, this could lead to mis-sequencing of cells/packets from the same VOQ. Mis-sequencing can be prevented very simply by

a) implementing the VOQs in RAM buffers instead of separate FIFO buffers and
   - re-ordering the time slots in which particular cells or packets within the same VOQ are transmitted from the real input ports (linecards), in the 1st stage; i.e. cells or packets would not exit a VOQ in the time order of their position in the VOQ (no longer FIFO)
   - not implementing the 1st-stage time-slot interchangers

OR

b) implementing the VOQs in FIFO buffers and
   - reading the cells or packets from their VOQs in FIFO order

- re-ordering the time slots in which the cells or packets are forwarded into the switch using 1st-stage time-slot interchangers

Path-searching (i.e. time-space channel assignment) is performed first through the middle-stage (4th) switches, by assigning specific middle-stage switches (channels) to each sub-group or sub-network. Secondly, each individual sub-group or sub-network is path-searched (time-slot assigned). Although path-searching algorithms for rearrangeably non-blocking switching networks could be employed, we will assume, for example, that just a single level of parallelism from the parallel path-search algorithm for 3-stage (Clos) networks from ~~IPD Case A25962~~ WO01/67802 is employed to assign channels and time slots. Port processors would operate in parallel, each one finding the required number of free, common channels between a pair of input and output ports, but the channels themselves would be found by sequential inspection.

First Path-Search Through Middle-Stage Switches

The first path-search through the middle-stage switches (assignment of LF time-space channels) would require N processors, each taking O(NLF) computing steps. Each of the processors must therefore have a clock speed $O(NLF/F\tau) = O(NL/\tau)$ Hz, which is 20 Gbit/s in our example. This is too high for today's sequential processors. Nevertheless, the ratio of processing capacity to switch capacity is

$$\frac{Processing Capacity}{Switch Capacity} = \frac{N}{LNB} O\left(\frac{NL}{\tau}\right) = O\left(\frac{N}{\tau B}\right) = O\left(\frac{N}{500}\right) \qquad \text{...eqtn. 4}$$

which is around a ratio of 0.064. The reduction from LN real ports to N sub-groups or sub-networks provides a proportional reduction in processing capacity compared with eqtn.3. The aggregation technique is therefore extremely promising, but another degree of parallelism is required to reduce the computing steps, without increasing the number of processors too much. To do this, let us employ n separate processors to perform sequential searching for free, common channels, each one searching sequentially through only O(LF/n) of the O(LF) channels. Each one would count the number of free channels that it finds. Then the individual counts would be added sequentially (e.g. by one of the n processors) in O(n) steps. When the sum reaches or exceeds the required number of channels, the adding can stop. The channels found by all the processors before the one that reached or exceeded the required number can be seized and assigned to both the relevant, paired sub-groups or sub-networks.

Alternatively, if they haven't yet been assigned, these processors can search again, assign and seize. The processor that reached or exceeded the required number can then search again, stopping when it has found, assigned and seized the last required channel. For this level of parallelism in the first path search:

$$\frac{Processing Capacity}{Switch Capacity} = \frac{\frac{NnN}{F\tau} O\left(\frac{LF}{n} + n + \frac{LF}{n}\right)}{LNB} = O\left(\frac{N}{\tau B}\left(2 + \frac{n^2}{LF}\right)\right) = O\left(\frac{N}{500}\left(2 + \frac{n^2}{LF}\right)\right) \quad \text{...eqtn. 5}$$

Because this ratio increases as the square of the number n of processors performing the sequential search in each port "processor", the number n should be as small as possible. Let us assume that a maximum processor clock speed of 2 GHz is available, hence

$$O\left(\frac{N}{F\tau}\left(\frac{LF}{n} + n + \frac{LF}{n}\right)\right) = 2 \quad \textbf{GHz} \quad \text{...eqtn. 6}$$

for which n is 21 in our example. This would lead to a processing/switching ratio

$$\frac{Processing Capacity}{Switch Capacity} = O\left(\frac{2.01N}{500}\right) = O\left(\frac{N}{250}\right) \quad \text{...eqtn. 7}$$

Thus, in our example, the penalty for reducing the processor speeds from 20 GHz to a practical value of 2 GHz is merely to double the processing capacity requirement. The resulting ratio for this first path search is just 0.128.

Second Path-Search Through Individual Sub-Networks

Path-searching (time-slot assignment) is next performed within each individual sub-group or sub-network. There are 2N sub-groups or sub-networks, each of which is path-searched in parallel. Furthermore, each sub-network is path-searched using L port processors in parallel, using the first level of parallelism of the parallel path-search algorithm for 3-stage Clos networks in ~~IPD Case A25962~~ WO 01/67802. Each one of these L port processors is assumed to search sequentially through O(F) time slots. (Of course blocking may require somewhere between F and 2F-1 time slots to be used). Hence the total number of processors is 2NL, each taking O(LF) computing steps. The processor clock speed required is O(LF/Fτ) = O(L/τ) = 640 Mbit/s, which is practicable. The processing/switching capacity ratio is

$$\frac{Processing Capacity}{Switch Capacity} = \frac{\left(\frac{2NL^2F}{F\tau}\right)}{LNB} = \frac{2L}{\tau B} = \frac{L}{250} \quad \text{...eqtn. 8}$$

which is the same value 0.128 in our example.

Thus from eqtns.7 and 8 the total ratio of processing capacity to switch capacity is 0.256. For a 10 Tbit/s switch this is an acceptable 2.56 Tbit/s.

Alternatively, the individual sub-groups or sub-networks could be path-searched using Andresen's algorithm (for example, see Steinar Andresen, "The looping algorithm extended to base 2^t rearrangeable switching networks," IEEE Trans. On Comms., Vol. COM-25, No. 10, 1057-1063 (1977), the text of which is hereby incorporated by reference), which uses a mapping between rearrangeably non-blocking Benes networks, which use multiple stages of 2x2 switches, and 3-stage Clos networks. This could result in an even lower processing capacity requirement. Andresen's algorithm can be implemented sequentially or using parallel processors. In practice, there is likely to be an optimum number of parallel processors that should be employed in order to reduce the processor speed to an acceptable value while achieving an acceptable total processing capacity. As this is a rearrangeably non-blocking algorithm, it guarantees that no more than F time slots are required within the sub-groups or sub-networks.

A Worked Example

To demonstrate the operation of the aggregated channel assignment technique in multi-stage switch fabrics, an example 4x4 cell or packet switch with L=2 ports per sub-group or sub-network, N=2 sub-networks, LN=4 ports (input and output) and frame duration F=4 time slots will be used. The traffic matrix for this example is taken from the results of the multi-stage matching algorithm in ~~IPD Case A30294~~ the co-pending application "MULTI-STAGE MATCHING PROCESS" by the inventors using the modified NOB25 pointer up-date rule which is deterministic for input ports but non-deterministic for output ports. The matrix of accepted requests for the next frame is

$$[a_2(i,j)] = \begin{bmatrix} 0 & 0 & 1 & 3 \\ 0 & 2 & 2 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{bmatrix} \qquad \text{...eqtn. 9}$$

Each matrix element $a_2(i,j)$ represents the number of cells or packets to be taken from each of the VOQs (i,j) at the 4 real input ports (linecards) and switched across the fabric in the next frame to the 4 real output ports (linecards). The logical 7-stage switch architecture shown in Figure 3 has 4 logical input and output ports for each

real input and output port, i.e. 16 logical input and output ports altogether. Each one of these represents an individual cell or packet to be switched across the fabric.

Assigning Cells or Packets to Input and Output Ports

Figure 3 shows schematically the result of assigning the 16 logical inputs and 16 logical outputs of the network to the cells or packets. This is not done for every cell or packet individually. Instead, on the input side of the switch, every real port or linecard i assigns a number $a_{i,j}$ of contiguous logical inputs (i.e. time slots) to the $a_{i,j}$ successfully accepted requests of its 4 VOQs. Similarly, on the output side, every real port or linecard j assigns a number $a_{i,j}$ of contiguous logical outputs to the $a_{i,j}$ successfully accepted requests from each of the 4 VOQs destined for it. The cells or packets are identified by three numbers. The first two give the identity of the cell's or packet's VOQ (i,j), the third number in brackets is the specific identity of the cell or packet in its VOQ. For convenience the position of the cell or packet within its VOQ is used as its specific identity. For example, the HOL packet in VOQ 1,3 is designated 1,3(1). The contiguous cells or packets in the same VOQ could be assigned to the logical input and output ports in any order. However, for convenience, they are assigned in the increasing order of their positions in their VOQ.

First Path-Search Through Middle-Stage Switches

The first path search requires N = 2 processors. The parallel path-search algorithm for 3-stage switches from ~~IPD Case A25962~~ WO 01/67802 requires the two left-hand-side sub-networks and the two right-hand-side sub-networks to be taken in pairs. Each pair is processed by one of the 2 processors. This can be performed simultaneously, in parallel. For example, pairs ($S_{lhs}1$, $S_{rhs}1$) and ($S_{lhs}2$, $S_{rhs}2$) could be processed first. When these have been path-searched, the pairings are changed and path-searching undertaken for the next set of sub-network pairs. For example, one of the sub-networks in the pairings could be incremented by 1 (cyclic pairing). Of course in the present switch example, there is only one other set of pairings possible; ($S_{lhs}1$, $S_{rhs}2$) and ($S_{lhs}2$, $S_{rhs}1$). Each processor adds up the total number of accepted requests in the VOQs between the pair of sub-networks it is dealing with. From eqtn.9 the reduced matrix of accepted request numbers between the sub-networks becomes

$$[a(S_{lhs}i, S_{rhs}j)] = \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \quad \text{..eqtn. 10}$$

Each processor then finds the first $a(S_{lhs}i, S_{rhs}j)$ free, common middle-stage switches (channels) that have not already been seized, between the first pair of sub-networks. The search begins from the uppermost, first middle-stage switch. This method is traditional "call packing". Each processor then proceeds to repeat the path-search for its second pairing of sub-networks. The assignment of a middle-stage switch to a particular pair of sub-networks also decides the path through the middle-stage switch. After all the accepted requests have been assigned a middle-stage switch (channel), the processors can assign specific cells or packets to those middle-stage switches. For convenience, cells or packets within the same VOQ are assigned to the middle-stage switches in the increasing order of their positions within the VOQ. The result of all this is to guarantee that the sequence of cells or packets within the same VOQ is preserved in the sequence of middle-stage switches (channels) assigned to them. The paths and cell/packet assignments after the first path-search phase through the middle-stage switches (channels) are shown in Figure 4. LF = 2x4 = 8 middle-stage switches (channels) are sufficient; none of the 16 "connections" is blocked.

Second Path-Search Through Individual Sub-Networks

Although the individual sub-networks could either be path-searched using parallel path-searching for 3-stage Clos networks or using a rearrangeably non-blocking algorithm such as the looping algorithm or Andresen's algorithm, the paths shown in Figure 5 have been established using the parallel path-search algorithm. This is performed in a similar fashion to the first path search.

To establish the paths, a reduced matrix of numbers of "connections" [c(S)] is first calculated between the 1st- and 3rd- stage switches of each sub-network from the cell/packet assignments on either side of the sub-network, as follows:

$$[c(S_{lhs}1)] = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{..eqtn. 11}$$

$$[c(S_{lhs}2)] = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{..eqtn. 12}$$

$$[c(S_{rhs}1)] = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \quad \text{..eqtn. 13}$$

$$[c(S_{rhs}2)] = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{..eqtn. 14}$$

These serve a similar role for each sub-network to that of eqtn.10 in the first path search. The identities of the specific cells or packets are then assigned to the middle-stage switches of each sub-network (i.e. the 2nd and 6th stages of the 7-stage logical network).

The cell/packet assignments are now sufficient throughout the network to be able to complete the paths through the 1st-, 3rd-, 5th- and 7th-stage switch matrices. Figure 6 of the accompanying drawings shows the complete paths assigned to every cell or packet. All cells or packets from the same VOQ are in sequence and contiguous at the logical outputs of the 7th stage, which means that they will be in sequence and contiguous when they are transmitted to line.

However, if the 7th logical stage of time-slot interchangers is left out, so that cells or packets can go straight out to line without having to be buffered at the real output ports (linecards), then cell/packet 3,2(1) and 3,2(2) will become out of sequence and lose contiguity, as will cells/packets 1,4(1), 1,4(2) and 1,4(3). Although loss of contiguity may be acceptable, mis-sequencing may not be. Fortunately, mis-sequencing can easily be prevented, without requiring any additional re-sequencing buffers. The method of prevention is purely algorithmic. At each 7th-stage logical switch, the order in which cells/packets within the same VOQ appear at the input ports of the switch can be used to re-order the time sequence in which the cells/packets must be forwarded from their VOQs in the 1st stage. For example, the cells/packets in VOQ 1,4 appear in the order 1,4(2), 1,4(3), 1,4(1) on the input ports of 7th-stage logical switch 4. The paths taken through the network by these cells/packets, which start at the logical output ports of the 1st-stage switch 1, must be mapped to cells/packets at the logical input ports of the 1st-stage switch 1 whose identities correspond to this order; i.e. 1,4(2)→1,4(1), 1,4(3)→1,4(2) and 1,4(1)→1,4(3). The connections across 1st-stage switch 1 are rearranged to provide this mapping, which of course re-orders the time sequence in which the cells/packets will be transmitted from the 1st-stage logical switch 1. Figure 7 shows the re-ordered

cell/packet path assignments for preventing mis-sequencing when the 7th-stage time-slot interchangers are left out; i.e. when there is no buffering in the real output ports (linecards) before transmission to line. Cells/packets 1,4(1), 1,4(2), 1,4(3), 3,2(1) and 3,2(2) have been re-ordered at the logical outputs of the 1st-stage time-slot interchangers 1 and 3.

Hence the HOL cell/packet 1,4(1) will be transmitted after cell/packet 1,4(3), which is behind it in VOQ 1,4, and before cell/packet 1,4(2). HOL cell/packet 3,2(1) is also transmitted after cell/packet 3,2(2) in 1st-stage logical switch 3. If the VOQs are implemented as FIFO buffers, then the 1st-stage switches must be implemented as time-slot interchangers so that the sequence of cells/packets can be re-ordered. But if the VOQs can be implemented as RAM buffers instead of FIFOs, then the re-ordering can be performed in the process of reading out the cells/packets from the RAM buffer, so that no additional time-slot interchangers are required in the 1st logical stage.. Evidently the 7th-stage switches must be included logically in the path-searching algorithm, in order to determine the correct sequence in which cells or packets must be forwarded from their VOQs to prevent mis-sequencing, but they are not physically implemented.

Those skilled in the art will appreciate that the number of stages can vary and may depend on the number of levels of aggregation used. For larger switch arrangements it is possible to have a larger number of buffer stages. The invention can be implemented in any suitable form, including software and/or hardware and the software algorithm may be provided in a form which is distributed amongst several components. The invention may be also implemented as a suite of one or more computer programs.

The invention may be implemented such that the egress subelements are not grouped into egress elements in one embodiment of the example, providing the total number of egress subelements equal the total number of ingress subelements. Those skilled in the art will appreciate that the invention can be implemented in a switching arrangement where the ingress elements are bidirectional and comprise said egress elements (and also where both ingress and egress elements are provided and are each bidirectional).

The text of the abstract is hereby deemed incorporated as part of the specification of the invention.

CLAIMS

1. A multi-stage time-slot assignment process for an input-queued switch arrangement in a communications network, the switch comprising a plurality of N ingress elements and N egress elements, each of the ingress elements having a number L of ingress subelements and each of the egress elements having a plurality L of egress subelements, the switch arrangement being arranged to have L or more real middle stage space switches of size N x N, and having F or more time-slots, the time-slot assignment process between the said ingress subelements and egress sub-elements comprising the steps of:

aggregating F time slots from each of a plurality L in number of said ingress subelements to form an ingress element having a plurality LF or more in number of time-space channels which are pooled between the L subelements of each ingress element and the L subelements of each egress element, wherein each time-space channel corresponds to a different logical middle-stage switch of the packet switch arrangement so that the number of logical input elements and logical output elements for which channel assignment is performed through the middle stage of the switch is N;

performing time-space channel assignment through the middle stage of the switch between the logical input elements and the logical output elements;

providing time-slot interchange capabilities at the 1st, 3rd, 5th stages; and

performing time-slot assignment between the ingress subelements of the ingress elements and the logical output ports of the 3rd stage switches, and performing time-slot interchange between the logical input ports of the 5th stage switches and the egress subelements of the egress elements.

2. A time-slot assignment process as claimed in claim 1, wherein the switch arrangement is arranged to switch cells or packets.

3. A time-slot assignment process as claimed in claim 1, wherein the switch arrangement is a circuit switch.

4. A time-slot assignment process as claimed in claim 2, wherein the switch arrangement is a cell switch arrangement capable of switching packets.

5. A time-slot assignment process as claimed in any preceding claim, further comprising a 7th stage providing a time-slot interchange capability.

6. A time-slot assignment process as claimed in any preceding claim, wherein the subelements comprise ports, and the elements comprise aggregations of ports.

7. A time-slot assignment process as claimed in any preceding claim, wherein the input-queued switch comprises VOQs which are implemented in random access memory RAM.

8. A time-slot assignment process as claimed in any preceding claim, further comprising recursive decomposition of the three stages in each element into seven stages using the steps indicated in claim 1.

9. A switch arrangement in which time-slots are assigned by a time-slot assignment process as claimed in any preceding claim.

10. A network having a switch arrangement as claimed in claim 9.

11. A suite of at least one computer program arranged to implement the time-slot assignment process according to any one of claims 1 to 8.

12. A suite of at least one computer program as claimed in claim 11, at least partly arranged to be implemented in hardware.

13. A suite of at least one computer program as claimed in claim 11 or 12, at least partly arranged to be implemented in software.

14. A time-slot assignment process substantially as described herein, with reference to the accompanying drawings.

ABSTRACT

MULTI-STAGE TIME-SLOT ASSIGNEMENT PROCESS

A multi-stage time-slot assignment process for an input-queued switch arrangement in a communications network, the switch comprising a plurality of N ingress elements and N egress elements, each of the ingress elements having a number L of ingress subelements and each of the egress elements having a plurality L of egress subelements, the switch arrangement being arranged to have L or more real middle stage space switches of size N x N, and having F or more time-slots, the time-slot assignment process between the said ingress subelements and egress subelements comprising the steps of: aggregating F time slots from each of a plurality L in number of said ingress subelements to form an ingress element having a plurality LF or more in number of time-space channels which are pooled between the L subelements of each ingress element and the L subelements of each egress element, wherein each time-space channel corresponds to a different logical middle-stage switch of the packet switch arrangement so that the number of logical input elements and logical output elements for which channel assignment is performed through the middle stage of the switch is N; performing time-space channel assignment through the middle stage of the switch between the logical input elements and the logical output elements; providing time-slot interchange capabilities at the 1st, 3rd, 5th stages; and performing time-slot assignment between the ingress subelements of the ingress elements and the logical output ports of the 3rd stage switches, and performing time-slot interchange between the logical input ports of the 5th stage switches and the egress subelements of the egress elements.

Figure (3)

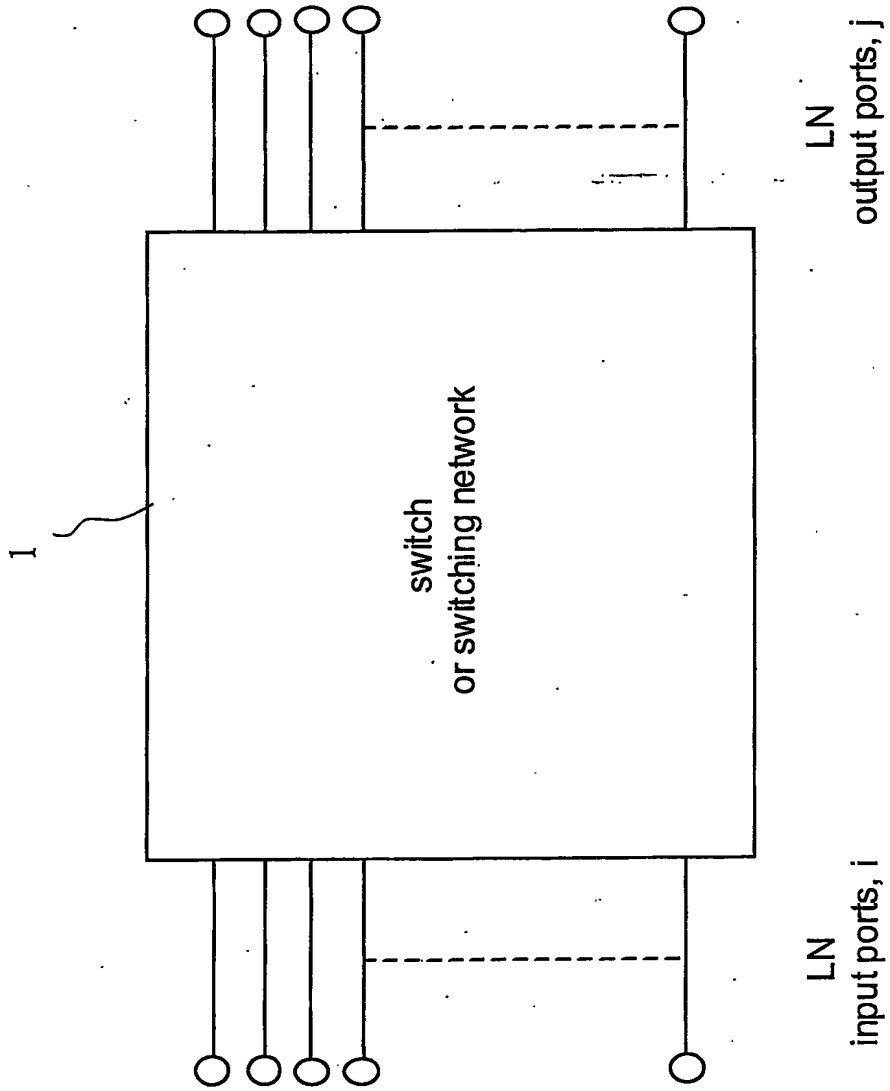

1
switch
or switching network
LN
output ports, j
LN
input ports, i

FIG. 1

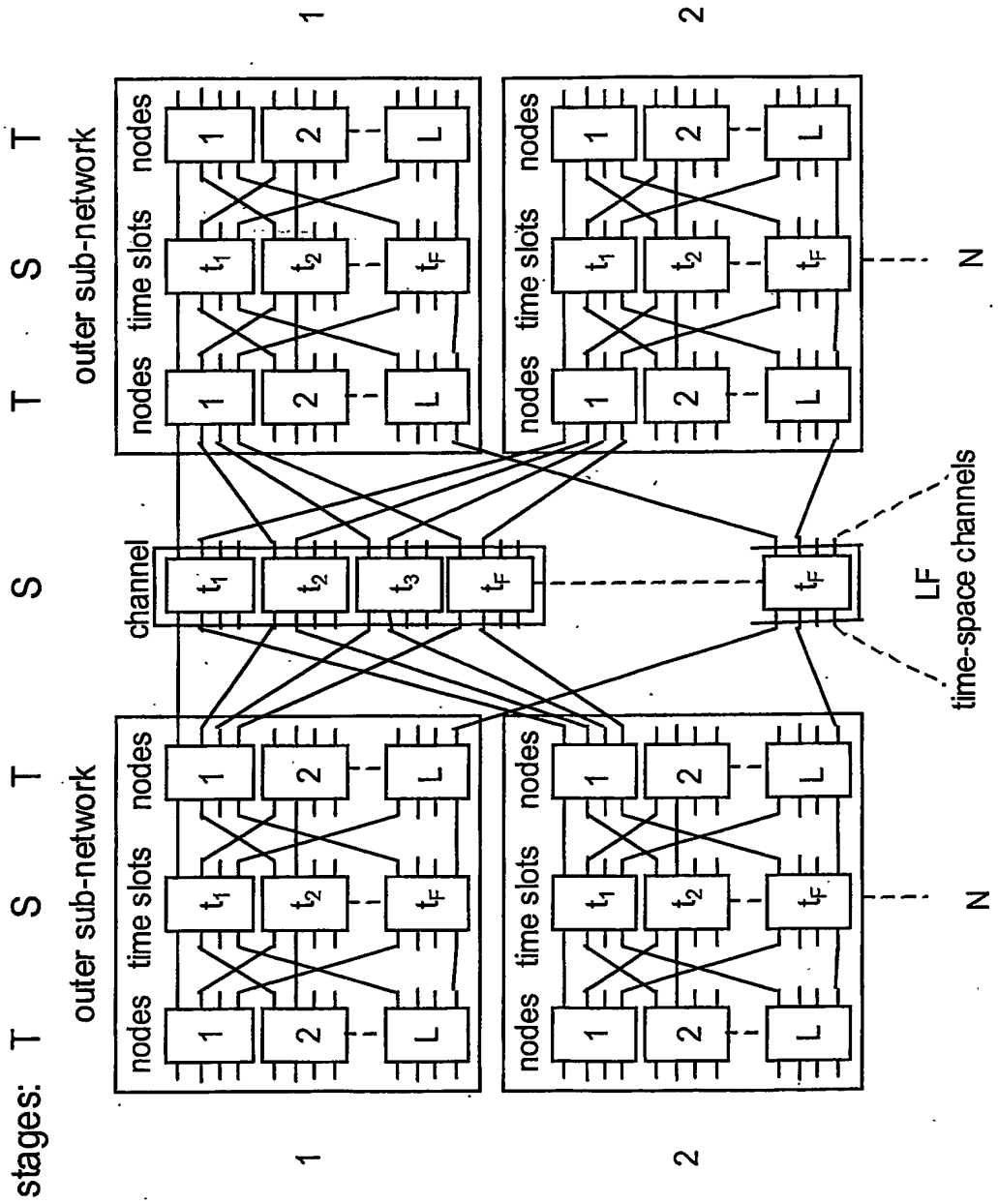

stages: T S T S T S T S T
outer sub-network
outer sub-network
nodes
time slots
nodes
channel
1
2
L
N
LF
time-space channels

FIG. 2

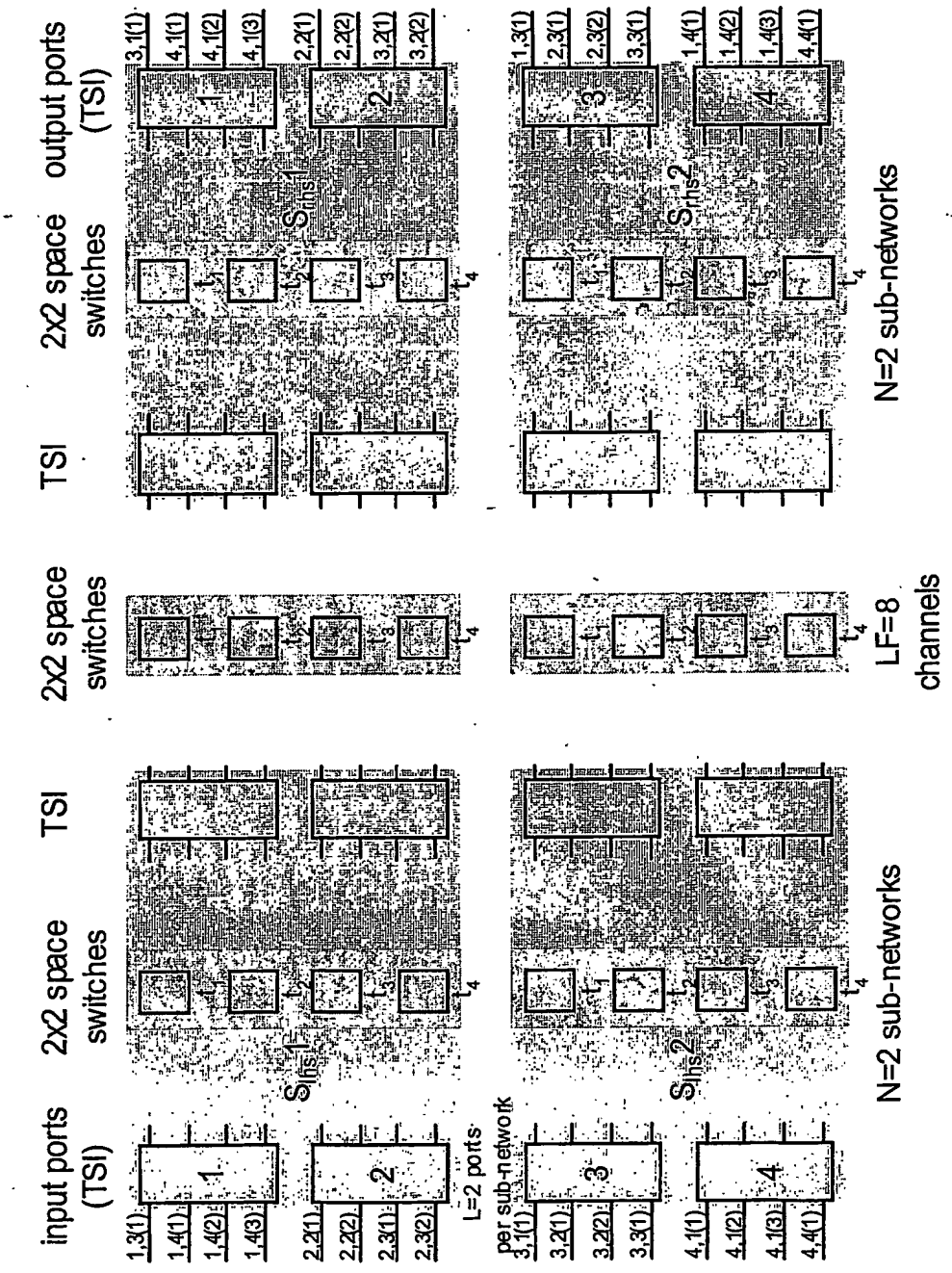

input ports (TSI)
2x2 space switches
TSI
2x2 space switches
TSI
output ports (TSI)
L=2 ports per sub-network
N=2 sub-networks
LF=8 channels
N=2 sub-networks

FIG. 3

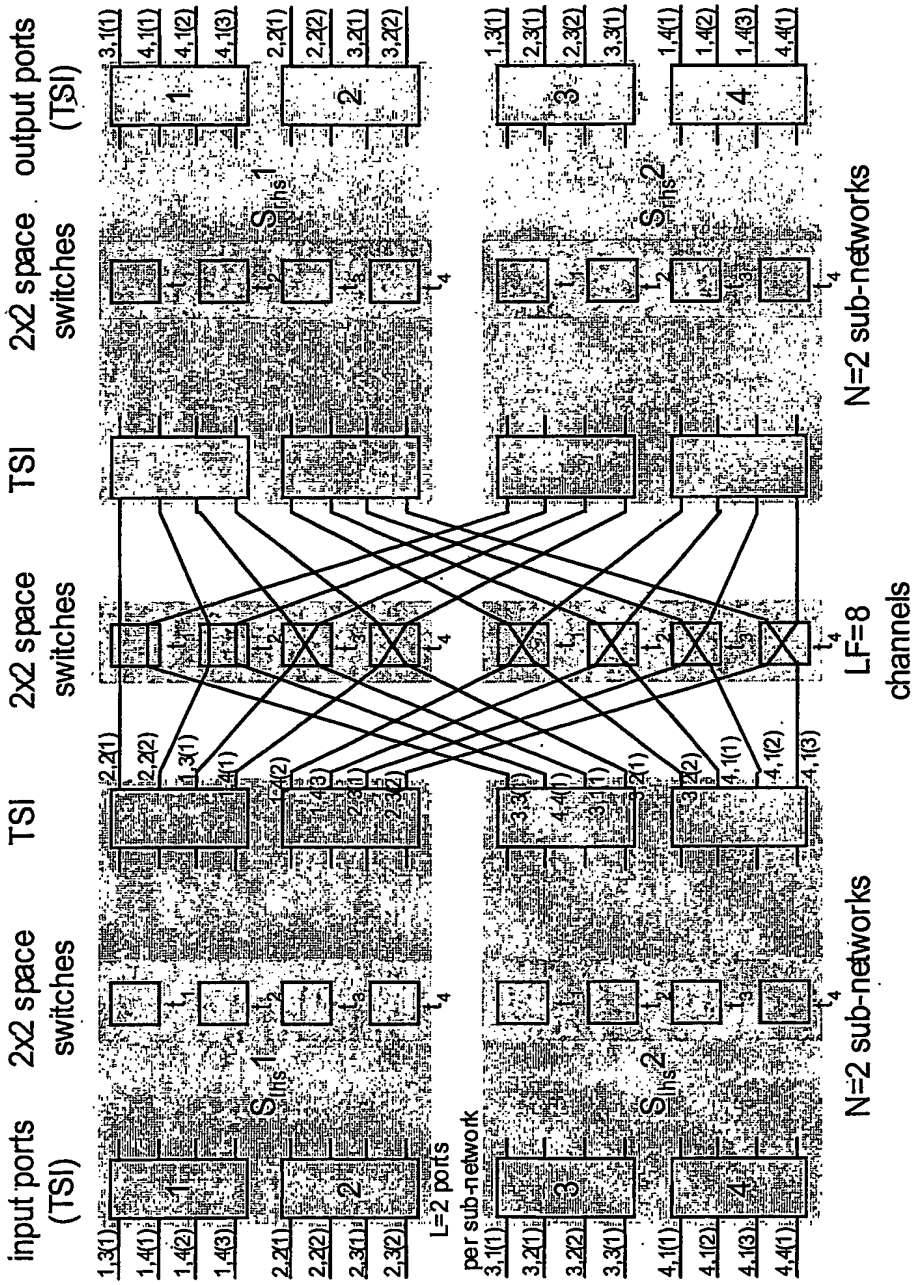

input ports (TSI)
2x2 space switches
TSI
2x2 space switches
TSI
2x2 space switches
output ports (TSI)
L=2 ports per sub-network
N=2 sub-networks
LF=8 channels
N=2 sub-networks

FIG. 4

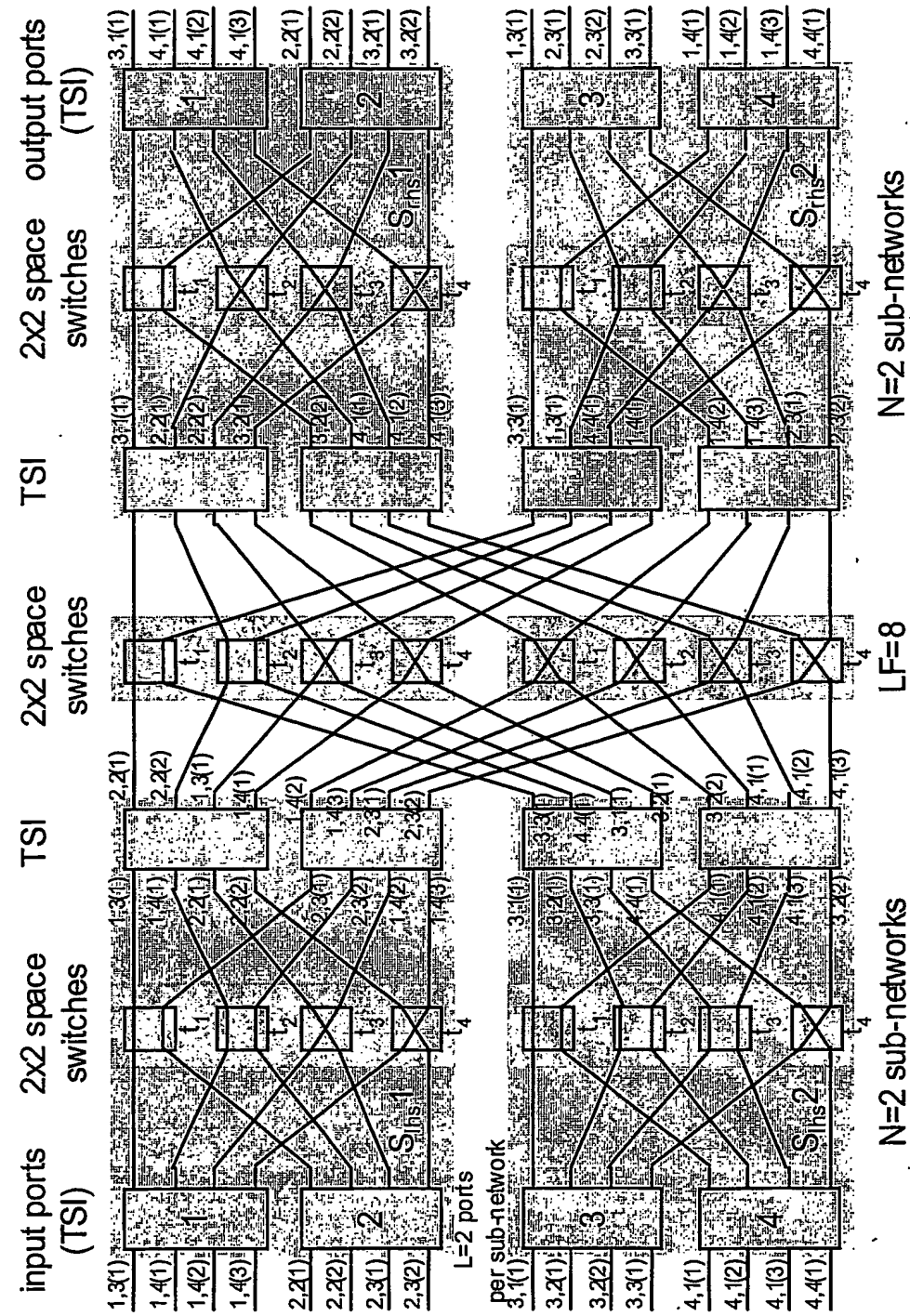

input ports (TSI)
2x2 space switches
TSI
2x2 space switches
TSI
2x2 space switches
output ports (TSI)
L=2 ports per sub-network
N=2 sub-networks
LF=8 channels
N=2 sub-networks

FIG. 5

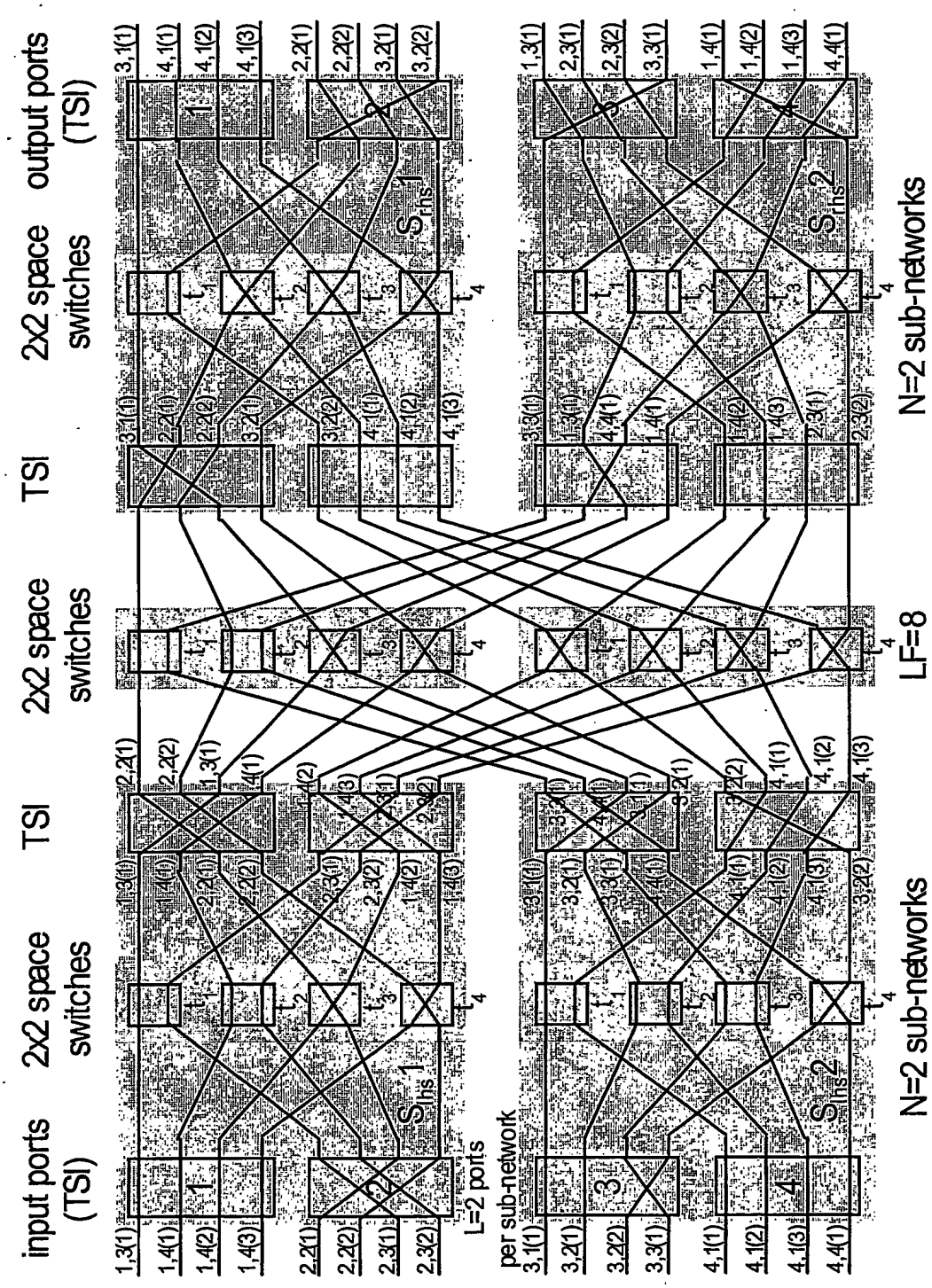

input ports (TSI)
2x2 space switches
TSI
2x2 space switches
TSI
2x2 space switches
output ports (TSI)
L=2 ports per sub-network
N=2 sub-networks
LF=8 channels
N=2 sub-networks

FIG. 6

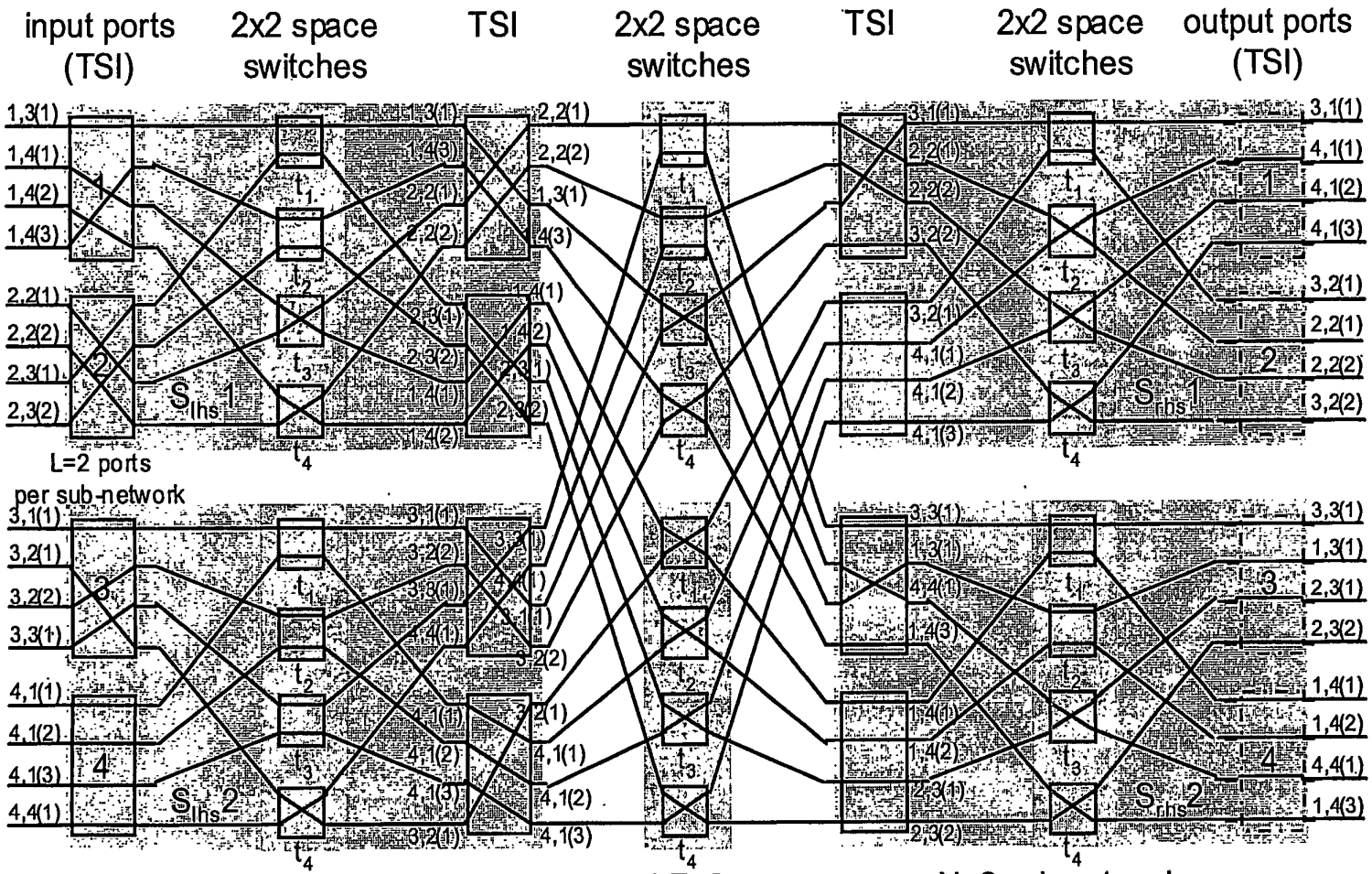

input ports (TSI)
2x2 space switches
TSI
2x2 space switches
TSI
2x2 space switches
output ports (TSI)
L=2 ports per sub-network
N=2 sub-networks
LF=8 channels
N=2 sub-networks

FIGURE 7

PCT/GB2004/004148